REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
xx-05-2012	Technical	May 2012

4. TITLE AND SUBTITLE
Evaluating DLAB as a Predictor of Foreign Language Learning

5a. CONTRACT NUMBER
H92222-10-D-0017/0007

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

SWA Consulting, Inc.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

SWA Consulting Inc.
311 S. Harrington St.
Suite 200
Raleigh, NC 27603

8. PERFORMING ORGANIZATION REPORT NUMBER

2012010609

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

HQ USSOCOM, Special Operations Forces Language Office
Attn: FMD-LDEMO-TL
7701 Tampa Point Blvd.
MacDill AFB, FL 33621-5323

10. SPONSOR/MONITOR'S ACRONYM(S)

SOFLO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

A. Approved for public release; distribution is unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This report evaluates the relationship between language learning aptitude and foreign language proficiency in the U.S. Army Special Operations Forces (ARSOF) community. The Defense Language Aptitude Battery (DLAB) is an assessment of language learning aptitude that is widely used in the ARSOF community for the purposes of selection and language training placement. This report includes two studies: *Study 1* documents the relationship between DLAB scores and proficiency attainment during initial acquisition training (IAT) at the United States Army John F. Kennedy Special Warfare Center and School (USAJFKSWCS); *Study 2* shows the relationship between DLAB scores and foreign language proficiency attainment over the duration of an ARSOF operator's career. The findings of Study 1 indicate DLAB scores are predictive of the level of speaking proficiency trainees attain during IAT. Trainees with higher DLAB scores were more likely to attain higher speaking proficiency than lower DLAB trainees. Study 2 findings indicate DLAB scores are predictive of the maximum proficiency ARSOF operators attain over the duration of their careers. Operators with higher DLAB scores tend to attain higher levels of proficiency than those with lower DLAB scores. These findings will help inform the use of DLAB scores for IAT language placement.

15. SUBJECT TERMS
ARSOF, DLAB, language learning aptitude, initial acquisition training, foreign language proficiency attainment

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Surface, Eric A.
U	U	U	UU (SAR)	32	**19b. TELEPHONE NUMBER** *(include area code)* 919-480-2751

Evaluating DLAB as a Predictor of Foreign Language Learning

SWA Consulting

MAY 2012

APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION UNLIMITED
SPONSORED BY: SOFLO, USSOCOM
RESEARCH CONDUCTED BY: SWA CONSULTING INC.

EXECUTIVE SUMMARY

Issue and Overview

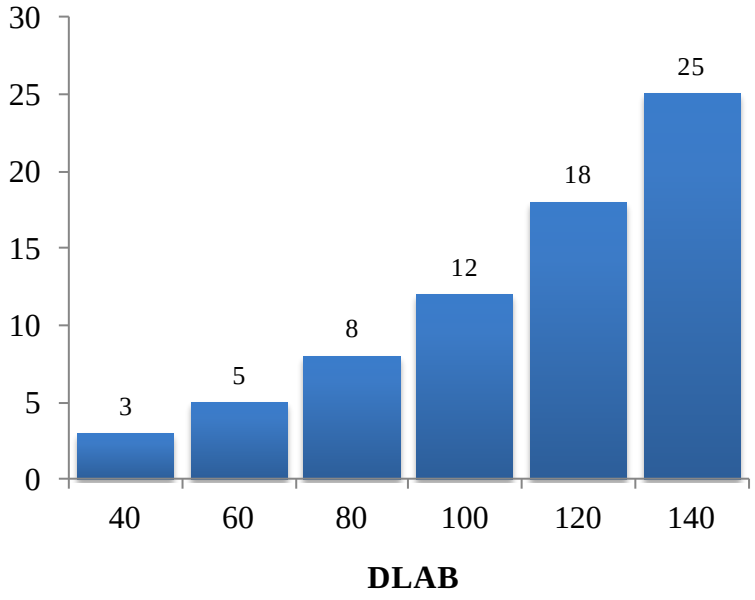
The purpose of this report is to evaluate the relationship between language learning aptitude and foreign language proficiency in the U.S. Army Special Operations Forces (ARSOF) community. The Defense Language Aptitude Battery (DLAB) is an assessment of language learning aptitude that is widely used in the ARSOF community for the purposes of selection and language training placement. Previous research has shown the DLAB to be one of the strongest single predictors of language learning during initial acquisition training (IAT) in the ARSOF community (e.g., SWA Consulting Inc., 2008, 2009). Therefore, this report focuses on the DLAB as the most readily available measure of language learning aptitude for ARSOF trainees.

Study 1 of this report documents the relationship between DLAB scores and proficiency attainment during IAT at the United States Army John F. Kennedy Special Warfare Center and School (USAJFKSWCS). *Study 2* shows the relationship between DLAB scores and foreign language proficiency attainment over the duration of an ARSOF operator's career. These findings will help inform the use of DLAB scores for IAT language placement.

Findings

Do DLAB scores predict which trainees are most likely to attain Interagency Language Roundtable (ILR) Level 2 speaking proficiency in a foreign language on the two-skill Oral Proficiency Interview (OPI) following IAT at USAJFKSWCS?

Number of Trainees (Out of 100) Expected to Attain ILR 2 Speaking Proficiency During USAJFKSWCS IAT[1]



[1] Values shown are averaged across the 13 training languages at USAJFKSWCS. Estimates will vary by language.

- *Study 1* findings indicate DLAB scores are predictive of the level of speaking proficiency trainees attain during IAT. Trainees with higher DLAB scores were more likely to attain higher speaking proficiency than lower DLAB trainees.
- The above figure illustrates the implication of this finding for trainees attaining ILR 2 speaking proficiency during IAT.
- **As shown in the figure above, trainees with higher DLAB scores attain ILR 2 speaking proficiency at a higher rate than those with lower DLAB scores.** For example, 18 out of 100 trainees with a DLAB score of 120 are expected to attain ILR 2, which is over two times the number expected to attain ILR 2 with a DLAB score of 80. While there were clear language differences in the overall likelihood of attaining ILR 2, the relationship of language learning aptitude to post-training proficiency held across languages.

Do DLAB scores predict the maximum foreign language proficiency level ARSOF operators attain over the duration of their careers?

- *Study 2* findings indicate DLAB scores are predictive of the maximum proficiency ARSOF operators attain over the duration of their careers. **Operators with higher DLAB scores tend to attain higher levels of proficiency than those with lower DLAB scores.**
- The following table summarizes this relationship, showing the estimated maximum proficiency level an operator is expected to attain for a "typical" language based on his DLAB score.

DLAB Score and Expected Maximum Proficiency Over the Duration of an ARSOF Operator's Career

DLAB Score	Expected Maximum Proficiency (ILR) for a "Typical" Language[2]				
	Speaking[a]	Listening		Reading	
		All DLPT Versions[b]	DLPT 5 Only[c]	All DLPT Versions[d]	DLPT 5 Only[e]
60	1	0+, 1	0+	1	0, 0+, 1
80	1	1	0+, 1	1	0, 0+, 1
100	1, 1+	1	1	1	1
120	1+	1	1	1, 1+, 2	1, 1+
140	1+, 2	1, 1+	1	1+, 2, 2+, 3	1, 1+, 2

[a] N = 1,657 across 43 languages. [b] N = 5,391 across 60 languages. [c] N = 4,757 across 44 languages.
[d] N = 5,363 across 61 languages. [e] N = 4,707 across 41 languages.

- A supplemental analysis is also provided in this report (Appendix C, p. 26), in which the relationship between DLAB scores and career foreign language proficiency attainment is evaluated for a more diverse sample of military personnel (including U.S. Army and U.S. Navy SOF and non-SOF).

[2] "Typical" refers to a language in which the average proficiency rating falls in the center of the distribution of ratings across all languages. Though language differences exist, these values best represent all of the languages for which test data is available.

- The relationship between DLAB scores and language proficiency in this diverse sample is generally consistent with that in the ARSOF sample.

Caveats

While the specific proficiency estimates presented in *Study 1* are helpful for visualizing the importance of language aptitude to foreign language learning in IAT, to apply them to current or future ARSOF trainees assumes no major changes in language training or testing from the historical sample. One notable change that has recently taken place at USAJFKSWCS is an increase in the training duration (and total contact hours) from 18 weeks to 24 weeks for Cat I/II languages. All Cat I/II trainees in the historical sample trained under the 18-week schedule. Since increased training duration should result in increased proficiency attainment, the specific proficiency estimates provided in this study will likely underestimate the proficiency Cat I/II trainees actually attain in a 24-week course. However, the primary goal of *Study 1* was to document the *relationship* between DLAB scores and proficiency attainment in IAT, which is not expected to change dramatically with the 30% increase in Cat I/II training duration.

The expected maximum proficiency levels presented in *Study 2* should be interpreted as the most likely proficiency level an ARSOF operator is expected to attain based solely on DLAB score using historical norms within the ARSOF community. Individuals with the same DLAB scores testing in the same language will certainly differ in the maximum proficiency levels they attain during their career. Also, two recent changes in ARSOF language testing are the introduction of the DLPT Version 5 (i.e., DLPT 5) for certain languages and the move from the DLPT Listening and Reading to the two-skill OPI (speaking and listening) as the test of record in 2008. We evaluated the influence of DLPT test version on the proficiency estimates, and present these findings in Appendix B. The change in proficiency standard from the DLPT to the two-skill OPI resulted in a shift in training emphasis towards speaking skills and away from reading skills. This shift in training emphasis may change the maximum proficiencies attained by operators in all language skill modalities. However, it is premature to evaluate the impact the change in testing standard will have on maximum career proficiency for any of the skill modalities examined in this report.

Conclusions and Recommendations

- DLAB scores are predictive of which trainees are most likely to attain ILR 2 speaking proficiency on the two-skill OPI following IAT. DLAB scores are also predictive of the maximum proficiency level ARSOF operators attain over the duration of their career.
  - ***Recommendation:*** **Special Warfare Education Group (Airborne)[3] leadership at USAJFKSWCS should continue to utilize the DLAB to inform language placement for IAT**.
- In an absolute sense, the relationship between DLAB scores and IAT proficiency ratings was small in magnitude. DLAB scores accounted for a fairly small percentage (less than 5%) of ARSOF trainees' proficiency attainment in IAT. The vast majority of the differences in actual proficiency attainment are due to factors other than language learning aptitude. These factors likely include trainee characteristics (other than language aptitude) and attitudes, experience, behaviors, and the characteristics of instruction and feedback trainees receive. Training

[3] Hereafter referred to as SWEG(A).

Diagnostic Reports (TDRs) and Instructor Feedback Reports (IFRs) provided by SWA Consulting to USAJFKSWCS throughout an IAT training course focus on these trainee- and class-level factors. Information provided in these reports allows SWEG(A) leadership to investigate and correct potential trainee and classroom issues before the conclusion of the training event. However, to be most effective, these resources need to be distributed to QASP personnel and language supervisors in a timely manner.

  - ***Recommendation*: SWEG(A) leadership should ensure that the Quality Assurance Surveillance Program (QASP) personnel and language supervisors at USAJFKSWCS receive Instructor Feedback Reports (IFRs) and Training Diagnostic Reports (TDRs) in a timely manner. Both IFRs and TDRs facilitate the provision of feedback from supervisors to instructors.**

- DLAB scores are currently used to inform not only selection into ARSOF, but also placement into specific training languages. The findings from *Study 1* and *Study 2* support the use of the DLAB for selection and language placement. However, language learning aptitude is one of many possible early indicators of successful language acquisition in the ARSOF context. Other early indicators might include general cognitive ability (e.g., Wonderlic Personnel Test™), educational attainment, age, prior language learning experience, motivation, and other trainee characteristics (e.g., personality, interests). These additional measures are collected during and immediately following Special Forces Assessment and Selection (SFAS), but are not used systematically to inform language assignment. More research is needed to evaluate whether or not these additional measures can be leveraged, either as supplements or alternatives to the DLAB, to optimize language assignment of ARSOF trainees. SWA Consulting is currently conducting this analysis, which is scheduled for release in Q3 2012.
  - ***Recommendation*: Future research should determine the most effective strategy for leveraging the DLAB and other cognitive and noncognitive measures assessed during SFAS to optimally assign ARSOF operators to training languages.**

This project was conducted by SWA Consulting Inc. under a subcontract with CACI-WGI, Inc. (Subcontract# B11-114482; Prime# H92222-10-D-0017/0007; Sub-CLIN 0003AB). For questions or more information about the SOFLO and this project, please contact Mr. Jack Donnelly (john.donnelly@socom.mil). For specific questions related to data collection or this report, please contact Dr. Eric A. Surface (esurface@swa-consulting.com) or Dr. Reanna Poncheri Harman (rpharman@swa-consulting.com).

CONTENTS

REPORT DETAILS

STUDY 1: DLAB AND ATTAINING ILR 2 SPEAKING PROFICIENCY DURING INITIAL ACQUISITION TRAINING

The focus of *Study 1* is to document the relationship between DLAB scores and proficiency attainment during initial acquisition training (IAT) at the United States Army John F. Kennedy Special Warfare Center and School (USAJFKSWCS). DLAB has been shown to be predictive of acquiring Interagency Language Roundtable (ILR) 1 and ILR 1+ proficiency during initial acquisition language training (IAT) (SWA Consulting Inc., 2008, 2009). However, the specific focus of *Study 1* is the attainment of ILR 2 speaking proficiency. *Study 1* addressed the following research question:

- Do DLAB scores predict which trainees are most likely to attain ILR 2 speaking proficiency on the Oral Proficiency Interview (OPI) following IAT?

Sample

The sample consisted of 1,814 trainees (from 492 classes) who completed foreign language IAT at USAJFKSWCS between 2009 and 2011. Training duration was 18 weeks for Category I and II languages and 24 weeks for Category III and IV languages. All training cohorts in the current sample followed the tri-semester schedule. Following training, trainees took the two-skill OPI (which tests speaking and participatory listening proficiency) to fulfill the graduation requirement of ILR 1 proficiency in two language modalities. In this sample, OPI speaking (OPI-S) and participatory listening (OPI-L) ratings were in exact agreement for over 99% of trainees. We focus only on OPI-S ratings, as the OPI-L ratings would exhibit virtually identical findings. DLAB records for trainees were obtained from the Defense Manpower Data Center (DMDC) and from organizational records from Special Forces Assessment and Selection (SFAS).

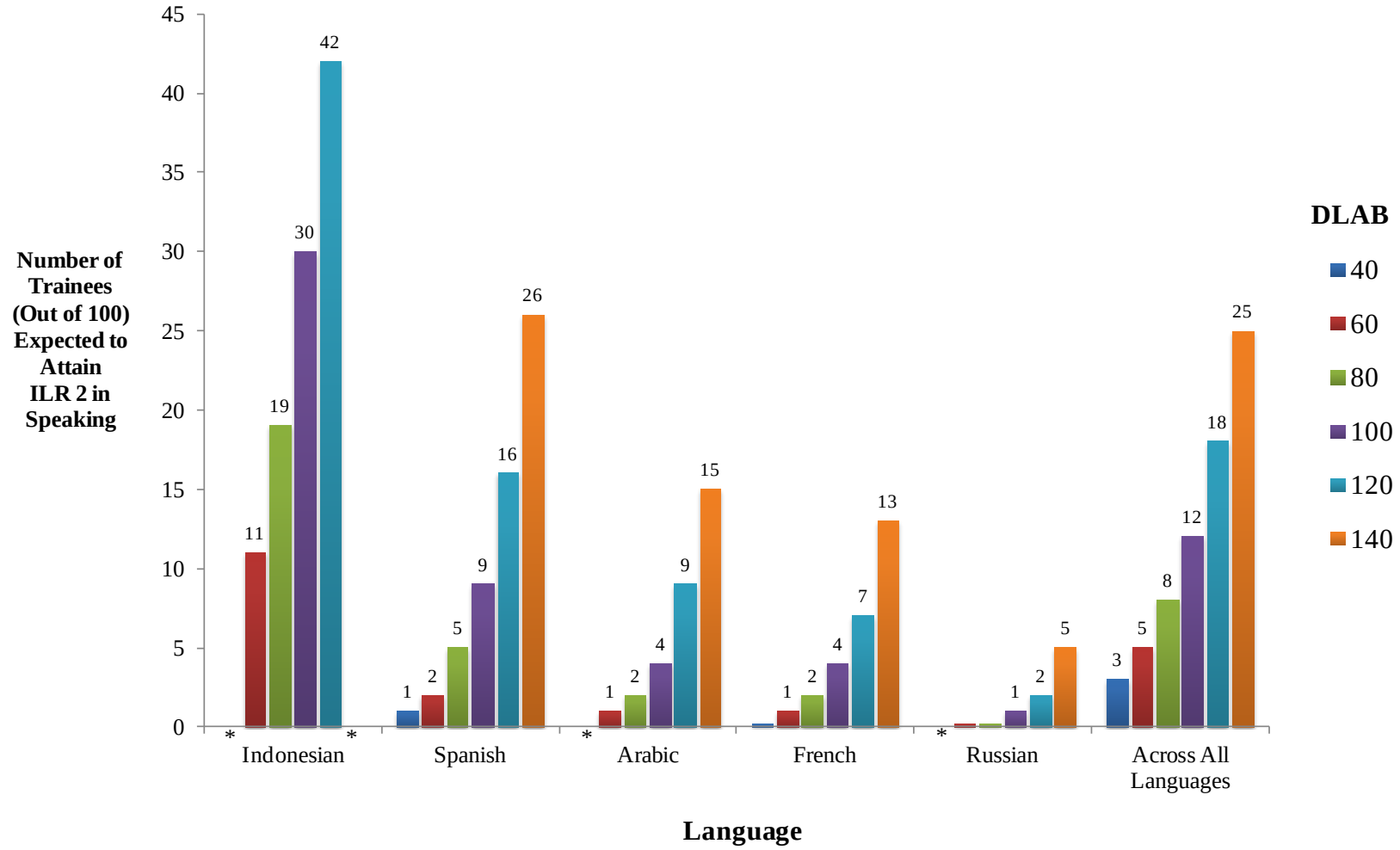
Findings

Results from this analysis revealed DLAB scores were predictive of trainees' OPI-S ratings. Details of this analysis are provided in Appendix A (p. 20). Trainees with higher DLAB scores attained higher OPI-S ratings across the full range of ILR ratings observed in this sample.[4] After accounting for training language, DLAB scores explained an additional 4.5% of the differences in trainees' OPI-S ratings. This finding is consistent with previous research showing DLAB scores accounting for approximately 3% of the differences in OPI-S ratings (SWA Consulting, 2009).

Figure 1 (p. 8) provides model-based estimates of the number of trainees (out of 100) expected to attain ILR 2 speaking proficiency by the completion of IAT at various levels of language learning aptitude. These estimates are provided for the five most commonly assigned languages at USAJFKSWCS. As an example, for trainees in Spanish, 16 out of 100 trainees with DLAB a score of 120 are expected to attain ILR 2. For Spanish trainees with a DLAB score of 80, only 5 out of 100 are expected to attain ILR 2. Figure 1 (p. 8) also provides estimates averaged across all 13 IAT languages at USAJFKSWCS (labeled "Across All Languages").

[4] OPI-S ratings ranged from ILR 0+ to ILR 2+ in the current sample. We cannot assume these findings extend to ILR levels higher than 2+ without additional data on trainees who attain those higher levels.

Figure 1. DLAB Scores and Attaining ILR 2 Speaking Proficiency During Training at USAJFKSWCS



Note. Languages presented in descending order of expected ILR 2 trainees. "*" indicates a DLAB score that was not observed in the target language. Projections are only provided for DLAB scores that were observed in the data. Select languages displayed here are the most commonly trained at USAJFKSWCS. "Across All Languages" category includes languages not shown.

There are two notable trends in Figure 1:

1. Across all 13 languages there is a clear relationship between DLAB scores and likelihood of attaining ILR 2. **An increase of 40 points on the DLAB from one trainee to another translates to an over 100% increase in the probability of attaining ILR 2 speaking proficiency during training**.

2. **A trainee's likelihood to attain ILR 2 is highly dependent on the language to which he is assigned.** Nearly 3 out of 10 trainees with a DLAB score of 100 are expected to attain ILR 2 in Indonesian, while only 1 out of 100 are expected to do so in Russian. While language differences are to be expected, these differences were not solely attributable to language category. For instance, trainees of all language aptitude levels were more likely to attain ILR 2 in Modern Standard Arabic (Cat IV) than Russian (Cat III). Similarly, holding language aptitude constant, more trainees attained ILR 2 in Indonesian (Cat II) compared to Spanish (Cat I) and French (Cat I). These findings suggest factors other than language category (e.g., trainee motivation, quality of instruction) influence proficiency attainment during IAT.

Caveat

While the specific proficiency estimates presented in this study are helpful for visualizing the importance of language aptitude to foreign language learning in IAT, to apply them to current or future ARSOF trainees assumes no major changes in language training or testing from the historical sample. One notable change that has recently taken place at USAJFKSWCS is an increase in the training duration (and total contact hours) from 18 weeks to 24 weeks for Cat I/II languages. All Cat I/II trainees in the historical sample trained under the 18-week schedule. Under the current training schedule, new trainees in Cat I/II languages will complete approximately 30% more training hours than trainees under the previous 18-week schedule. Since increased training duration should result in increased proficiency attainment, the specific proficiency estimates provided in this study will likely underestimate the proficiency Cat I/II trainees actually attain in a 24-week course. However, the primary goal of this study was to document the *relationship* between DLAB scores and proficiency attainment in IAT, which is not expected to change dramatically with the 30% increase in Cat I/II training duration.

Conclusions

The objective of *Study 1* was to document the relationship between language learning aptitude (as measured by the DLAB) and attainment of ILR 2 speaking proficiency during USAJFKSWCS IAT. The findings indicate DLAB scores are in fact predictive of SOF trainees' attaining ILR 2 speaking proficiency during IAT. These findings add to those from prior studies (e.g., SWA Consulting Inc., 2008, 2009) by focusing on a level of proficiency (i.e., ILR 2 and greater) that is historically high for ARSOF trainees completing IAT.

These findings support the use of DLAB scores for placement into training languages, as trainees with greater language learning aptitude are shown to attain higher levels of proficiency across training languages than lower aptitude trainees. DLAB scores provide a valid indicator of learning potential in the USAJFKSWCS IAT training context, and should continue to inform selection and language placement decisions.

While these findings support the relationship between DLAB scores and ILR speaking proficiency, it is important to note the magnitude of this relationship was small to moderate in magnitude. This indicates that, within a given language, the differences in trainees' OPI ratings are largely due to factors other than language learning aptitude. These factors, among others, likely include trainees' attitudes towards training (motivation to learn), learning goals, study behaviors (time-on-task outside of the classroom), and the quality of instruction and feedback trainees receive. Other individual and situational factors should be systematically investigated to better understand which trainees achieve ILR 2.

STUDY 2: DLAB AND MAXIMUM FOREIGN LANGUAGE PROFICIENCY ATTAINED BY ARSOF OPERATORS

Study 1 of this report focused on foreign language proficiency attained during the Qualification Course required to become an ARSOF operator. However, language skills gained during the Qualification Course represent only an initial snapshot of the broader progression of skills an operator will exhibit throughout his career. In an effort to better understand what trainee characteristics contribute to long-term sustainment and development of language proficiency, this study examines how language learning aptitude relates to the maximum level of proficiency individuals attain during their careers. Specifically, *Study 2* addressed the following research question:

- Do DLAB scores predict the maximum proficiency level ARSOF operators attain over the duration of their career?

Sample

An archival sample was obtained using testing records—including DLPT and OPI ratings—and DLAB scores provided by the DMDC. The initial pool of records included DLPT ratings for over 25,500 individuals and OPI ratings for over 8,500 individuals who either were members of SOF or participated in training through SOF between 1995 and 2011. Of this initial pool, only ARSOF operators (as determined by Military Occupational Specialty [MOS]) were retained in for the final sample.[5] The final sample selected for inclusion in the sample varied by skill modality, with 1,657 (across 43 languages) included in the speaking proficiency analysis, 5,391 (across 60 languages) included in the listening proficiency[6] analysis, and 5,363 (across 61 languages) included in the reading proficiency analysis. MOSs for the individuals in the sample are presented in the Appendix B (Table B1, p. 21). The testing languages included in these analyses are presented in Appendix B (Table B2, p. 22-23). For cases in which an individual tested in multiple languages over the course of his career, his maximum proficiency for each language was included in the analysis.

Findings

Speaking Proficiency

The findings indicated DLAB scores were predictive of the maximum speaking proficiency attained during an ARSOF operator's career. Details of this analysis are provided in the Appendix B (p. 23). Individuals with higher DLAB scores tended to attain higher levels of proficiency throughout their careers than those with lower DLAB scores. Holding the language in which one tests constant, 8.3% of the differences (or variability) in maximum proficiency attained by ARSOF operators was accounted for by DLAB score. Regarding the practical significance of this finding, Table 1 (p. 12) shows the expected maximum speaking proficiency rating for individuals with varying DLAB scores across the five most commonly tested languages in this sample.

There are three notable trends in Table 1 (p. 12):

- **Individuals with higher DLAB scores were more likely to attain ILR 2 level speaking proficiency during their career than those with lower DLAB scores**.

[5] An analysis including individuals outside of ARSOF is presented in Appendix C (p. 26).
[6] All listening proficiency ratings were based on the DLPT. No OPI-L ratings were included.

- **Even individuals with relatively low DLAB scores (i.e., 60) attained ILR 1 speaking proficiency, regardless of the target language.** Maximum speaking proficiency ratings ranged from ILR 1 to ILR 2 for the majority of the sample.
- **An individual's likelihood to attain ILR 2 is somewhat dependent on the target language.**

Table 1. DLAB Score and Expected Maximum Attained Speaking Proficiency

DLAB Score	Expected Maximum Speaking Proficiency (ILR)					
	Spanish	French	Russian	Chinese-Mandarin	Arabic	*"Typical" Language* [a]
60	1	1	1	- [b]	1	*1*
80	1, 1+	1	1	1	1	*1*
100	1+	1, 1+	1	1	1	*1, 1+*
120	1+, 2	1+	1+	1+	1, 1+	*1+*
140	2	1+, 2	1+	1+	1+	*1+, 2*

N = 1,657 across 43 languages.

[a] "Typical" refers to a language in which the average proficiency rating falls in the center of the distribution of ratings across all languages. Though language differences exist, these values best represent all of the languages in the dataset. Includes languages not shown.

[b] This DLAB score was not observed in the target language. Projections are only provided for DLAB scores that were observed in the data.

Listening Proficiency

DLAB scores were predictive of the maximum listening proficiency ARSOF operators attained during their careers. Higher DLAB scores were associated with higher levels of maximum proficiency than those with lower DLAB scores. After accounting for the target language, DLAB scores explained 4.3% of the differences in listening proficiency ratings between operators in this sample. Table 2 (p. 13) illustrates the relationship between DLAB scores and maximum listening proficiency rating by indicating the most probable listening proficiency rating to be attained across a range of DLAB scores.

There are four notable trends in Table 2 (p. 13):

- **Individuals with higher DLAB scores were more likely to attain ILR 2 and ILR 3 level listening proficiency during their career than those with lower DLAB scores**.
- **Individuals with relatively low DLAB scores (i.e., 60) tended to reach their maximum listening proficiency at ILR 0+ or 1 across languages.** Maximum listening proficiency ratings ranged from ILR 0+ to ILR 3 for the majority of the sample.
- **DLAB scores did not allow for highly precise predictions of maximum listening proficiency.** This lack of precision is evident in Table 2 (p. 13). In Spanish, for example, an individual with a DLAB score of 100 was equally likely to attain proficiency levels between ILR 1 and ILR 3. This finding indicates factors other than language learning aptitude substantially influence the maximum listening proficiency level attained by ARSOF operators.

- **An individual's likelihood to attain ILR 2 or ILR 3 is dependent in part on the target language.** For example, individuals with moderate to high DLAB scores were much more likely to attain ILR 3 in Spanish (Cat I) than in Chinese-Mandarin or Arabic (Cat IV).

Table 2. DLAB Score and Expected Maximum Attained Listening Proficiency (All DLPT Versions)

DLAB Score	Expected Maximum Listening Proficiency (ILR)					
	Spanish	French	Russian	Chinese-Mandarin	Arabic	*"Typical" Language* [a]
60	1	1	1	-[b]	0, 0+	*0+, 1*
80	1, 1+, 2	1	1	1	0+, 1	*1*
100	1, 1+, 2, 2+, 3	1	1	1	1	*1*
120	2, 2+, 3	1, 1+, 2	1, 1+	1	1	*1*
140	3	1+, 2, 2+, 3	1, 1+, 2	1, 1+	1	*1, 1+*

N = 5,391 across 60 languages.

[a] "Typical" refers to a language in which the average proficiency rating falls in the center of the distribution of ratings across all languages. Though language differences exist, these values best represent all of the languages in the dataset. Includes languages not shown.

[b] This DLAB score was not observed in the target language. Projections are only provided for DLAB scores that were observed in the data.

Reading Proficiency

Similar to listening proficiency, DLAB scores were predictive of the maximum reading proficiency ARSOF operators attained during their careers. Higher DLAB scores were associated with higher levels of reading proficiency. Accounting for target language, DLAB scores explained 7.3% of the differences in reading proficiency ratings between operators in this sample. Table 3 (p. 14) shows the relationship between DLAB scores and maximum reading proficiency rating.

The trends in Table 3 mirror those for listening proficiency (Table 2, p. 13), with the following exception:

- **Individuals with fairly high DLAB scores (e.g., 120-140) were more likely to attain ILR 2 and ILR 3 proficiency in reading than in listening.** In general, reading proficiency ratings were slightly higher than listening proficiency ratings.

Table 3. DLAB Score and Expected Maximum Attained Reading Proficiency (All DLPT Versions)

DLAB Score	Expected Maximum Listening Proficiency (ILR)					
	Spanish	French	Russian	Chinese-Mandarin	Arabic	*"Typical" Language* [a]
60	1+, 2, 2+, 3	1	1	- [b]	0, 0+, 1	*1*
80	2, 2+, 3	1, 1+, 2	1	1	1	*1*
100	3	1, 1+, 2	1, 1+, 2	1, 1+	1	*1*
120	3	2, 2+, 3	1, 1+, 2	1, 1+, 2	1, 1+	*1, 1+, 2*
140	3	3	2, 2+, 3	1+, 2, 2+, 3	1, 1+, 2, 2+	*1+, 2, 2+, 3*

N = 5,363 across 61 languages.

[a] "Typical" refers to a language in which the average proficiency rating falls in the center of the distribution of ratings across all languages. Though language differences exist, these values best represent all of the languages in the dataset. Includes languages not shown.

[b] This DLAB score was not observed in the target language. Projections are only provided for DLAB scores that were observed in the data.

Caveats

The expected maximum proficiency levels presented in *Study 2* should be interpreted as the most likely proficiency level an ARSOF operator is expected to attain based solely on DLAB score using historical norms within the ARSOF community. Individuals with the same DLAB scores testing in the same language will certainly differ in the maximum proficiency levels they attain during their career. This fact is clear from the finding that DLAB scores account for less than 10% of the differences in proficiency between individuals testing in the same language. The expected maximum proficiency levels presented here should not be interpreted as the maximum *possible* proficiency individuals could attain given unlimited time, resources, and immersion opportunities. The nature of this study (e.g., archival, correlational) does not support that specific interpretation.

While the specific proficiency estimates presented in this study are helpful for visualizing the importance of language aptitude to maximum proficiency, applying them to current or future ARSOF operators assumes no major changes in language training or testing from the historical sample. Two notable and recent changes in ARSOF language testing are the introduction of the DLPT Version 5 (i.e., DLPT 5) for certain languages and the move from the DLPT Listening and Reading to the two-skill OPI (speaking and listening) as the test of record in 2008. We evaluated the influence of DLPT test version on the proficiency estimates, and present these findings in the Appendix B. The change in proficiency standard from the DLPT to the two-skill OPI resulted in a shift in training emphasis towards speaking skills and away from reading skills. This shift in training emphasis may change the maximum proficiencies attained by operators in all language skill modalities. However, it is premature to evaluate the impact the change in testing standard will have on maximum career proficiency for any of the skill modalities examined in this report.

As noted in *Study 1*, IAT training duration for Cat I/II languages was recently increased at USAJFKSWCS from 18 weeks to 24 weeks. This increase in training time for Cat I/II languages is likely

to influence the maximum proficiency those training in these languages attain during IAT, and ultimately, their careers in SOF. Additionally, there are other situational factors (e.g., organizational context, testing context) that may influence the maximum proficiency an ARSOF operator will attain that are not explicitly examined in this study. However, these variations in the organizational context, testing version, proficiency standard and training duration are not expected to dramatically affect the underlying relationship between language learning aptitude and the acquisition of proficiency, which is the focus of this report. While testing and training practices may alter the levels of proficiency operators attain across skill modalities, those higher in language learning aptitude are still expected to attain greater proficiency than lower aptitude individuals in a comparable testing and training environment.

Conclusions

The objective of *Study 2* was to establish the relationship between DLAB scores and the maximum proficiency level ARSOF operators attain over the duration of their career. The findings from this analysis show that DLAB scores are in fact predictive of the maximum proficiency operators ultimately attain.

It is important to note that the relationships between DLAB scores and proficiency ratings observed in this study were small to moderate in magnitude. For a given target language, DLAB only predicted between 4% (for listening) and 8% (for speaking) of the differences in operators' maximum proficiency on record. This means DLAB scores failed to predict as much as 92%-96% of the differences in maximum proficiency. These differences among operators may be due to other systemic factors, such as access and usage of learning resources, opportunity to practice and use the target language, participation in language immersion, and other individual trainee characteristics.

OVERALL CONCLUSIONS AND RECOMMENDATIONS

Foreign language proficiency is a critical job requirement for SOF operators. SOF operators frequently deploy on missions in which effective performance and mission success depend on the ability to communicate using the local language(s) (*2004 SOF Language Transformation Strategy Needs Assessment Project* [Technical Report #20040606], *2009 SOF Language and Culture Needs Assessment Project* [Technical Reports #2010011010 & #2009010618]). Therefore, language learning aptitude is considered to be an important attribute of an effective SOF operator. The purpose of this report is to evaluate the relationship between language learning aptitude (as assessed by the DLAB) and attainment of foreign language proficiency in the Special Operations Forces (SOF) community. This section provides recommendations based on the findings of this report.

Recommendation: SWEG(A) leadership should continue to utilize the DLAB to inform language placement for IAT at USAJFKSWCS. Findings from *Study 1* indicate DLAB scores are predictive of SOF trainees' attaining ILR 2 speaking proficiency during IAT. *Study 2* findings suggest language learning aptitude is predictive of the maximum proficiency SOF operators attain during their career. As would be expected, language learning aptitude appears to enable individuals to effectively learn and produce language at relatively high levels. This finding supports the validity of the DLAB as an indicator of potential for language learning success in the context of SOF foreign language training. Therefore, we recommend the leadership at the SWEG(A) continue to utilize the DLAB to inform language placement for IAT. Future studies should evaluate the relationship between language learning aptitude and proficiency attainment during the Intermediate Language Course (ILC) at USAJFKSWCS. Such a study would help evaluate the usefulness of DLAB scores in the ILC selection and placement process.

Recommendation: Ensure that the Quality Assurance Surveillance Program (QASP) personnel and language supervisors at USAJFKSWCS receive Instructor Feedback Reports (IFRs) and Training Diagnostic Reports (TDRs) in a timely manner. Perhaps less intuitive, however, is the finding that DLAB scores accounted for a fairly small proportion of SOF trainees' proficiency attainment in IAT (less than 5%). Comparing trainees assigned to the same training language, the vast majority of the differences in actual proficiency attainment is due to factors other than language learning aptitude. These factors likely include trainee attributes and attitudes, experience, behaviors, and the quality of instruction and feedback trainees receive. Training Diagnostic Reports (TDRs) and Instructor Feedback Reports (IFRs) provided by SWA Consulting to USAJFKSWCS throughout an IAT training course focus on these trainee- and class-level factors. The primary goal of these reports is to provide the SWEG(A) leadership with diagnostic feedback on all IAT classes as each cohort completes training. Both TDRs and IFRs facilitate the provision of feedback from supervisors to instructors. Additionally, these reports facilitate the prioritization and necessity of classroom observations. Information provided in these reports allows SWEG(A) leadership to investigate and correct potential trainee and classroom issues before the conclusion of the training event. However, to be most effective, these resources need to be distributed to QASP personnel and language supervisors in a timely manner.

Recommendation: Future research should determine the most effective strategy for leveraging the DLAB and other cognitive and noncognitive measures assessed during SFAS to optimally assign ARSOF operators to training languages. DLAB scores are currently used to inform not only selection

into ARSOF, but also placement into specific training languages. The findings from *Study 1* and *Study 2* support the use of the DLAB for selection and language placement. However, language learning aptitude is one of many possible early indicators of successful language acquisition in the ARSOF context. Other early indicators might include general cognitive ability (e.g., Wonderlic Personnel Test™), educational attainment, age, prior language learning experience, motivation, and other trainee characteristics (e.g., personality, interests). These additional measures are collected during and immediately following SFAS, but are not used systematically to inform language assignment. More research is needed to evaluate whether or not these additional measures can be leveraged, either as supplements or alternatives to the DLAB, to optimize language assignment of ARSOF trainees. SWA Consulting is currently conducting this analysis, which is scheduled for release in Q3 2012.

REFERENCES

Asparouhov, T. & Muthén, B. (2010a). *Bayesian analysis using Mplus: Technical appendix*. Los Angeles: Muthén & Muthén. www.statmodel.com

Muthén, L.K. and Muthén, B.O. (1998-2010). *Mplus User's Guide: Sixth Edition*. Los Angeles, CA: Muthén & Muthén

Spiegelhalter, D.J., Thomas, A., Best, N. & Gilks, W.R. (1997). *BUGS: Bayesian Inference Using Gibbs Sampling*, Version 0.60. Cambridge: Medical Research Council Biostatistics Unit

Surface, E. A., Poncheri, R., Lemmond, G., & Shetye, T. (2005, March). *Special Operations Forces language transformation strategy needs assessment project: Final project report*. (Technical Report #20040606). Raleigh, NC: Surface, Ward & Associates.

SWA Consulting Inc. (2008, November). *Wonderlic Personnel Test (WPT™) as a Predictor of Language Proficiency* (Technical Report #2008010608). Raleigh, NC: Author.

SWA Consulting Inc. (2009, June). *Predictors of proficiency on the OPI: Considering the WPT™, Army GT, AFQT, DLAB* (Technical Report #2009010612). Raleigh, NC: Author.

SWA Consulting Inc. (2010a). *2009 SOF language and culture needs assessment project: Inside AOR use of language report*. (Technical Report #2010011010). Raleigh, NC: Author.

SWA Consulting Inc. (2010b). *2009 SOF language and culture needs assessment project: Outside AOR use of language report*. (Technical Report #2009010618). Raleigh, NC: Author.

ABOUT SWA CONSULTING INC.

SWA Consulting Inc. (formerly Surface, Ward, and Associates) provides analytics and evidence-based solutions for clients using the principles and methods of industrial/organizational (I/O) psychology. Since 1997, SWA has advised and assisted corporate, non-profit and governmental clients on:

- Training and development
- Performance measurement and management
- Organizational effectiveness
- Test development and validation
- Program/training evaluation
- Work/job analysis
- Needs assessment
- Selection system design
- Study and analysis related to human capital issues
- Metric development and data collection
- Advanced data analysis

One specific practice area is analytics, research, and consulting on foreign language and culture in work contexts. In this area, SWA has conducted numerous projects, including language assessment validation and psychometric research; evaluations of language training, training tools, and job aids; language and culture focused needs assessments and job analysis; and advanced analysis of language research data.

Based in Raleigh, NC, and led by Drs. Eric A. Surface and Stephen J. Ward, SWA now employs close to twenty I/O professionals at the masters and PhD levels. SWA professionals are committed to providing clients the best data and analysis upon which to make evidence-based decisions. Taking a scientist-practitioner perspective, SWA professionals conduct model-based, evidence-driven research and consulting to provide the best answers and solutions to enhance our clients' mission and business objectives. SWA has competencies in measurement, data collection, analytics, data modeling, systematic reviews, validation, and evaluation.

For more information about SWA, our projects, and our capabilities, please visit our website (www.swa-consulting.com) or contact Dr. Eric A. Surface (esurface@swa-consulting.com) or Dr. Stephen J. Ward (sward@swa-consulting.com).

APPENDIX A: STUDY 1

Description of Analysis

To address the *Study 1* research question, a two-level regression model for a categorical outcome was conducted using M*plus* (Muthén & Muthén, 1998-2010). Level 1 represented the trainee (or within class) level of analysis, and Level 2 represented the class (or between class) level of analysis. The focus of this study was the trainee level of analysis (i.e., Level 1). For this model, OPI-S rating was specified as the categorical outcome variable, which varied between trainees. DLAB score was entered as the only Level 1 predictor variable. DLAB score was group-mean centered to remove bias in the estimation of the slope due to between-class variability.[7] The following Level 2 class-level covariates were entered as statistical controls: training language (dummy-coded), cohort, and average DLAB score for the class. Bayesian estimation was used to estimate model parameters (see Asparouhov & Muthen [2010] for technical implementation of Bayesian estimation in *M*plus). Uninformative priors were used for all parameters, with the exception of the random intercept variance term. As is typically done in Bayesian multilevel modeling (e.g., Spiegelhalter, Thomas, Best, & Gilks, 1997), a small variance prior using the inverse-gamma distribution (slope = .001, scale = .001) was used for the random intercept variance term. We used 1,000 iterations as burn-in and an additional 1,000 iterations for estimation. To reduce autocorrelation of estimates between iterations, a thinning factor of 10 was used, such that only estimates from every 10th iteration were retained for the posterior distributions.

The results indicated proper convergence for the model. Potential scale reduction (see Asparouhov & Muthen, 2010) achieved a sufficiently small value (1.003), indicating proper convergence. Also, autocorrelations between parameter estimates for adjacent iterations were low.

Table A1 (p. 20) presents the parameter estimates for DLAB scores predicting OPI-S ratings for the Level 1 (within class) portion of the analysis.[8] The results indicate a significant positive relationship between DLAB score and trainees' OPI-S rating.

Table A1. Within Class Results from Two-Level Regression of OPI-S Proficiency Rating on DLAB Score

	Unstandardized Estimate	Posterior S.D.	*p* (one-tailed)	95% Credibility Interval		Standardized Estimate
				Lower 2.5%	Upper 2.5%	
DLAB score	.02	< .01	< .001	0.01	0.02	0.21

Note. DLAB slope uses probit link function. Within class R^2 = .045, p < .001. n = 1814 trainees, 492 classes.

[7] Note that mean DLAB score was entered at Level 2. Therefore, between-class variability in the DLAB-OPI slope was represented at the appropriate level.
[8] Results for the between class portion are available upon request.

APPENDIX B: STUDY 2

Table B1. Military Occupational Specialties for Study Sample

MOS	Speaking Proficiency Sample		Listening Proficiency Sample		Reading Proficiency Sample	
180	20	1.2%	182	3.4%	182	3.4%
183	-	-	1	.0%	1	.0%
18A	262	15.8%	1431	26.5%	1427	26.6%
18B	50	3%	257	4.8%	254	4.7%
18C	50	3%	302	5.6%	300	5.6%
18D	35	2.1%	180	3.3%	178	3.3%
18E	54	3.3%	285	5.3%	284	5.3%
18F	9	0.5%	36	.7%	36	.7%
18X	777	46.9%	1395	25.9%	1383	25.8%
18Z	6	0.4%	118	2.2%	118	2.2%
37A	-	-	1	.0%	1	.0%
37F	293	17.7%	830	15.4%	828	15.4%
37X	-	-	-	-	1	.0%
38A	101	6.1%	372	6.9%	370	6.9%
Total	*1,657*	*100%*	*5,391*	*100%*	*5,363*	*100%*

Table B2. Languages Included in *Study 2* Analyses

Speaking Proficiency Analysis:

AA - Afrikaans
AB - Albanian
AD - Arabic (Modern Standard)
AE - Arabic-Egyptian
AP - Arabic-Syrian
CC - Chinese-Cantonese
CM - Chinese-Mandarin
CX - Czech
DG - Arabic-Gulf
FR - French
GM - German
GR - Greek
HC - Haitian-Creole
HE - Hebrew
HJ - Hindi
JA - Japanese
JN - Indonesian
JT - Italian
KP - Korean
ML - Malay
NE - Nepalese
NR - Norwegian
PF - Persian-Iranian
PG - Persian-Afghan
PJ - Punjabi
PL - Polish
PQ - Portuguese-Brazilian
PT - Portuguese-European
PU - Pashtu
PV - Pashtu-Afghan
PY - Portuguese
QB - Spanish
RU - Russian
SC - Serbo-Croatian
SK - Slovak
TA - Tagalog
TC - Tamil
TH - Thai
TU - Turkish
UK - Ukrainian
UR - Urdu
VN - Vietnamese-Hanoi
VY – Visayan/Bisayan

Listening Proficiency Analysis:

AB - Albanian
AD - Arabic (Modern Standard)
AE - Arabic-Egyptian
AP - Arabic-Syrian
AZ - Arabic
BU - Bulgarian
CA - Cambodian
CC - Chinese-Cantonese
CM - Chinese-Mandarin
CX - Czech
DA - Danish
DG - Arabic-Gulf
DU - Dutch
FE - Frisian
FJ - Finnish
FR - French
GE - Greek (New Testament)
GM - German
GR - Greek
GT - German-Bavarian
HC - Haitian-Creole
HE - Hebrew
HJ - Hindi
HS - Hausa
HU - Hungarian
JA - Japanese
JN - Indonesian
JT - Italian
KP - Korean
LA - Spanish-American
LC - Lao
LT - Lithuanian
NR - Norwegian
PF - Persian-Iranian
PG - Persian-Afghan
PJ - Punjabi
PL - Polish
PQ - Portuguese-Brazilian
PT - Portuguese-European
PU - Pashtu
PV - Pashtu-Afghan
PY - Portuguese
QB - Spanish
RQ - Romanian
RU - Russian
SA - Samoan
SC - Serbo-Croatian
SL - Slovenian
SP - Sotho
SR - Spanish-Castilian
SW - Swahili
SY - Swedish
TA - Tagalog
TH - Thai
TU - Turkish
UK - Ukrainian
UR - Urdu
VN - Vietnamese-Hanoi
XS - Sorani
YA - Yakut

Table B2 (continued). Languages Included in *Study 2* Analyses

Reading Proficiency Analysis:

AB - Albanian
AD - Arabic (Modern Standard)
AE - Arabic-Egyptian
AP - Arabic-Syrian
AR - Armenian
AZ - Arabic
BU - Bulgarian
CA - Cambodian
CC - Chinese-Cantonese
CM - Chinese-Mandarin
CX - Czech
DA - Danish
DG - Arabic-Gulf
DU - Dutch
FE - Frisian
FJ - Finnish
FR - French
GE - Greek (New Testament)
GM - German
GR - Greek
GT - German-Bavarian
HC - Haitian-Creole
HE - Hebrew
HJ - Hindi
HS - Hausa
HU - Hungarian
JA - Japanese
JN - Indonesian
JT - Italian
KP - Korean
LA - Spanish-American
LC - Lao
LT - Lithuanian
NR - Norwegian
PF - Persian-Iranian
PG - Persian-Afghan
PJ - Punjabi
PL - Polish
PQ - Portuguese-Brazilian
PT - Portuguese-European
PU - Pashtu
PV - Pashtu-Afghan
PY - Portuguese
QB - Spanish
RQ - Romanian
RU - Russian
SA - Samoan
SC - Serbo-Croatian
SL - Slovenian
SP - Sotho
SR - Spanish-Castilian
SW - Swahili
SY - Swedish
TA - Tagalog
TH - Thai
TU - Turkish
UK - Ukrainian
UR - Urdu
VN - Vietnamese-Hanoi
XS - Sorani
YA - Yakut

Description of Analysis

To address the *Study 2* research question, a two-level regression model for a categorical outcome was conducted using M*plus* (Muthén & Muthén, 1998-2010). Level 1 represented the individual (or within language) level of analysis, and Level 2 represented the language (or between language) level of analysis. The focus of this study was the individual level of analysis (i.e., Level 1). A separate model was estimated for each of the three proficiency outcomes of interest (i.e., OPI-S, DLPT-L, DLPT-R). For each model, proficiency ratings were specified as a categorical outcome variable, which varied between individuals. Maximum likelihood estimation, with standard errors that were robust to non-normality, was used. DLAB score was entered as the only Level 1 predictor variable. DLAB score was grand-mean centered. For the DLPT-R outcome, there was evidence of a small curvilinear relationship between DLAB scores and proficiency. Therefore, both a linear and quadratic DLAB slope were included at Level 1 for this outcome. No Level 2 covariates were included in these models. A random intercept (for the proficiency outcome) was estimated for each language to account for the nesting of individuals within testing language.

Table B3 (p. 24) presents the parameter estimates for DLAB scores predicting proficiency ratings for the Level 1 (within language) portion of the analyses.[9] The results indicate a significant positive relationship between DLAB score and individuals' proficiency ratings for all outcomes.

[9] Results for the between language portion are available upon request.

Table B3. Within Language Results from Two-Level Regression of Proficiency Ratings on DLAB Score

	Unstandardized Estimate	S.E.	*p* (two-tailed)	95% Confidence Interval		Standardized Estimate
				Lower 2.5%	Upper 2.5%	
Speaking Proficiency[a]						
DLAB score						
Linear slope	.03	< .01	< .001	.03	.04	.29
Listening Proficiency[b]						
DLAB score						
Linear slope	.02	< .01	< .001	.02	.02	.21
Reading Proficiency[c]						
DLAB score						
Linear slope	.03	< .01	< .001	.02	.03	.27
Quadratic slope	< .01	< .01	< .01	< .01	< .01	.03

Note. DLAB slope uses logit link function.

[a] Within language $R^2 = .08$, $p < .001$. n = 1,657 across 43 languages.

[b] Within language $R^2 = .04$, $p < .001$. n = 5,391 across 60 languages. Includes all DLPT versions.

[c] Within language $R^2 = .07$, $p < .001$. n = 5,363 across 61 languages. Includes all DLPT versions.

Analysis of DLPT Version and Maximum Proficiency on Record

Version 5 of the DLPT was rolled out for select languages in 2007-2008. To account for potential differences in proficiency ratings due to the rollout of the DLPT 5, we examined the extent to which DLPT test version influenced the maximum DLPT-L and DLPT-R proficiency rating on record for this sample. Specifically, we focused on the maximum proficiency level ARSOF operators would be expected to attain if testing on the DLPT 5.

For listening proficiency, the findings showed ratings on the DLPT 5 were lower than those on earlier DLPT versions. Accounting for DLPT version[10] explained an additional 2.9% of the differences in maximum listening proficiency ratings (controlling for DLAB scores). Table B4 (p. 25) shows the maximum listening proficiency ratings SOF operators are expected to attain assuming they are testing on the DLPT 5.

There is one notable trend in Table B4 in comparison to the prior *Study 2* analysis, which did not account for DLPT version (Table 2, p. 13):

- **Individuals tended to attain lower levels of listening proficiency on DLPT 5 compared to earlier DLPT versions.** As a result, the maximum proficiency level on record for an individual is expected to be somewhat lower if the rating was produced by a DLPT 5 rather than by an earlier DLPT version for a given language.

[10] We specifically compared DLPT 5 to prior versions. No comparisons between earlier versions were performed.

Table B4. DLAB Score and Expected Maximum Attained Listening Proficiency (DLPT 5 Only)

DLAB Score	Expected Maximum Listening Proficiency (ILR)					
	Spanish	French	Russian	Chinese-Mandarin	Arabic	*"Typical" Language* [a]
60	1	0+, 1	0+, 1	- [b]	0	*0+*
80	1	1	1	0+, 1	0, 0+	*0+, 1*
100	1, 1+	1	1	1	0+	*1*
120	1, 1+, 2	1	1	1	0+, 1	*1*
140	1, 1+, 2, 2+, 3	1, 1+	1, 1+	1	1	*1*

N = 4,757 across 44 languages.

[a] "Typical" refers to a language in which the average proficiency rating falls in the center of the distribution of ratings across all languages. Though language differences exist, these values best represent all of the languages in the dataset. Includes languages not shown.

[b] This DLAB score was not observed in the target language. Projections are only provided for DLAB scores that were observed in the data.

The findings for reading proficiency were similar to those for listening proficiency. Reading proficiency ratings on the DLPT 5 were lower than those on earlier DLPT versions. Controlling for DLAB scores, DLPT version explained an additional 3.7% of the differences in maximum reading proficiency ratings. Table B5 (p. 25) shows the maximum reading proficiency ratings SOF operators are expected to attain assuming they are testing on the DLPT 5. As with listening proficiency, the maximum reading proficiency level on record is expected to be somewhat lower if the rating was produced by a DLPT 5 rather than by an earlier DLPT version.

Table B5. DLAB Score and Expected Maximum Attained Reading Proficiency (DLPT 5 Only)

DLAB Score	Expected Maximum Listening Proficiency (ILR)					
	Spanish	French	Russian	Chinese-Mandarin	Arabic	*"Typical" Language* [a]
60	1, 1+, 2	0, 0+, 1	0, 0+, 1	- [b]	0	*0, 0+, 1*
80	1+, 2	1	1	1	0	*0, 0+, 1*
100	1+, 2, 2+, 3	1	1	1	0, 0+, 1	*1*
120	3	1, 1+, 2	1, 1+	1, 1+	1	*1, 1+*
140	3	1+, 2	1+, 2	1, 1+, 2	1, 1+	*1, 1+, 2*

N = 4,707 across 41 languages.

[a] "Typical" refers to a language in which the average proficiency rating falls in the center of the distribution of ratings across all languages. Though language differences exist, these values best represent all of the languages in the dataset. Includes languages not shown.

[b] This DLAB score was not observed in the target language. Projections are only provided for DLAB scores that were observed in the data.

APPENDIX C: DLAB AND MAXIMUM FOREIGN LANGUAGE PROFICIENCY ATTAINED IN A BROAD MILITARY SAMPLE

Study 2 of this report examined how language learning aptitude relates to the maximum level of proficiency ARSOF operators attain during their careers. In the following supplemental analysis, we examine this relationship in a broader military sample. This sample includes SOF and non-SOF personnel. This supplemental analysis addressed the following research question:

- Do DLAB scores predict the maximum proficiency level military personnel in general attain over the duration of their career?

Sample

An archival sample was obtained using testing records—including DLPT and OPI ratings—and DLAB scores provided by the DMDC. The initial pool of records included over DLPT ratings for over 25,500 individuals and OPI ratings for over 8,500 individuals who either were members of SOF or participated in training through SOF between 1995 and 2011. No specific MOSs were excluded from the final pool of records. A description of MOSs represented in this sample is provided in Table C1 (p. 27). The sample size varied by skill modality, with 3,475 (across 52 languages) included in the speaking proficiency analysis, 10,024 (across 71 languages) included in the listening proficiency[11] analysis, and 9,944 (across 70 languages) included in the reading proficiency analysis. For cases in which an individual tested in multiple languages over the course of his career, his maximum proficiency for each language was included in the analysis.

Findings

Speaking Proficiency

The findings indicated DLAB scores were predictive of the maximum speaking proficiency attained during an individual's career. This analysis (and those that follow in this section) followed the same procedure as that described in the Appendix B (p. 23). Individuals with higher DLAB scores tended to attain higher levels of proficiency throughout their careers than those with lower DLAB scores. Holding the language in which one tests constant, 11% of the differences (or variability) in maximum proficiency attained by individuals was accounted for by DLAB score. Table C2 (p. 27) shows the expected maximum speaking proficiency rating for individuals with varying DLAB scores across the five most commonly tested languages in this sample.

There are three notable trends in Table C2 (p. 27):

- **Individuals with higher DLAB scores were more likely to attain ILR 2 level speaking proficiency during their career than those with lower DLAB scores**.
- **Even individuals with relatively low DLAB scores (i.e., 40) attained ILR 1 speaking proficiency, regardless of the target language.** Maximum speaking proficiency ratings ranged from ILR 1 to ILR 2 for the majority of the sample.
- **An individual's likelihood to attain ILR 2 is somewhat dependent on the target language.**

[11] All listening proficiency ratings were based on the DLPT. No OPI-L ratings were included.

Table C1. Military Occupational Specialties for Study Sample

MOS	Speaking Proficiency Sample		Listening Proficiency Sample		Reading Proficiency Sample	
U.S. Army						
11 series	239	6.9%	670	6.7%	666	6.7%
18 series	1,243	35.8%	4,005	40.0%	3981	40.0%
180	20	0.6%	182	1.8%	182	1.8%
35 series	56	1.6%	264	2.6%	264	2.7%
37 series	293	8.4%	832	8.3%	830	8.3%
38 series	101	2.9%	372	3.7%	370	3.7%
97 series	250	7.2%	439	4.4%	440	4.4%
98 series	329	9.5%	691	6.9%	688	6.9%
Other	848	24.4%	2,329	23.2%	2298	23.1%
U.S. Navy	96	2.8	240	2.4%	225	2.3%
Total	*3,475*	*100%*	*10,024*	*100%*	*9,944*	*100%*

Note. Only the largest MOS categories in the sample are shown.

Table C2. DLAB Score and Expected Maximum Attained Speaking Proficiency

DLAB Score	Expected Maximum Speaking Proficiency (ILR)					
	Spanish	French	Russian	Chinese-Mandarin	Arabic	*"Typical" Language* [a]
40	1	1	1	1	1	*1*
60	1	1	1	1	1	*1*
80	1+	1	1	1	1	*1, 1+*
100	1+, 2	1+	1, 1+	1+	1, 1+	*1+*
120	2	1+, 2	1+	1+, 2	1+	*2*
140	2	2	2	2	2	*2*

N = 3,475 across 52 languages.

[a] "Typical" refers to a language in which the average proficiency rating falls in the center of the distribution of ratings across all languages. Though language differences exist, these values best represent all of the languages in the dataset. Includes languages not shown.

Listening Proficiency

DLAB scores were also predictive of the maximum listening proficiency individuals attained during their careers. Higher DLAB scores were associated with higher levels of maximum proficiency than those with lower DLAB scores. After accounting for the target language, DLAB scores explained 5% of the differences in listening proficiency ratings between operators in this sample. Table C3 (p. 28) illustrates the relationship between DLAB scores and maximum listening proficiency rating.

There are four notable trends in Table C3 (p. 28):

- **Individuals with higher DLAB scores were more likely to attain ILR 2 and ILR 3 level listening proficiency during their career than those with lower DLAB scores**.

- **Individuals with relatively low DLAB scores (i.e., 40-60) tended to reach their maximum listening proficiency at ILR 0+ or 1 across languages.** Maximum listening proficiency ratings ranged from ILR 0+ to ILR 3 for the majority of the sample.
- **DLAB scores did not allow for highly precise predictions of maximum listening proficiency.** This lack of precision is evident in Table C3 (p. 28). In French, for example, an individual with a DLAB score of 120 was equally likely to attain proficiency levels between ILR 1 and ILR 3. This finding indicates factors other than language learning aptitude substantially influence the maximum proficiency level attained by individuals in the sample.
- **An individual's likelihood to attain ILR 2 or ILR 3 is dependent in part on the target language.** For example, individuals with moderate DLAB scores were much more likely to attain ILR 3 in Spanish (Cat I) than in Chinese-Mandarin or Arabic (Cat IV).

Table C3. DLAB Score and Expected Maximum Attained Listening Proficiency (All DLPT Versions)

DLAB Score	Expected Maximum Listening Proficiency (ILR)					
	Spanish	French	Russian	Chinese-Mandarin	Arabic	*"Typical" Language* [a]
40	1	1	1	0+, 1	0, 0+, 1	*0+, 1*
60	1	1	1	1	0+, 1	*1*
80	1, 1+, 2, 2+, 3	1	1	1	0+, 1	*1*
100	3	1, 1+, 2	1, 1+	1	1	*1*
120	3	1, 1+, 2, 2+, 3	1, 1+, 2, 2+, 3	1, 1+, 2	1	*1*
140	3	3	3	2, 2+, 3	1, 1+, 2, 2+, 3	*1+, 2, 2+, 3*
160	3	3	3	3	2, 2+, 3	*3*

N = 10,024 across 71 languages.

[a] "Typical" refers to a language in which the average proficiency rating falls in the center of the distribution of ratings across all languages. Though language differences exist, these values best represent all of the languages in the dataset. Includes languages not shown.

Reading Proficiency

Similar to listening proficiency, DLAB scores were predictive of the maximum reading proficiency individuals attained during their careers. Higher DLAB scores were associated with higher levels of reading proficiency. Accounting for target language, DLAB scores explained 8.4% of the differences in reading proficiency ratings between individuals in this sample. Table C4 (p. 29) shows the relationship between DLAB scores and maximum reading proficiency rating.

The trends in Table C4 mirror those for listening proficiency (Table C3, p. 28), with the following exceptions:

- **DLAB scores provided slightly more precise estimates of maximum reading proficiency** than was the case for listening proficiency. In French, for example, an individual with a DLAB score of 100 was equally likely to attain reading proficiency between ILR 1+ and ILR 3 (a slightly smaller range than that seen for listening proficiency).
- **Individuals with moderate DLAB scores (e.g., 100-120) were more likely to attain ILR 2 and ILR 3 proficiency in reading than in listening.** In general, reading proficiency ratings were slightly higher than listening proficiency ratings.

Table C4. DLAB Score and Expected Maximum Attained Reading Proficiency (All DLPT Versions)

DLAB Score	Expected Maximum Listening Proficiency (ILR)					
	Spanish	French	Russian	Chinese-Mandarin	Arabic	*"Typical" Language* [a]
40	1, 1+, 2, 2+, 3	1	0, 0+, 1	1	0	*0, 0+, 1*
60	2, 2+, 3	1, 1+	1	1	0, 0+, 1	*1*
80	3	1, 1+, 2	1, 1+	1, 1+	1	*1*
100	3	1+, 2, 2+, 3	1, 1+, 2	1, 1+, 2, 2+	1	*1, 1+, 2*
120	3	3	1+, 2, 2+, 3	2, 2+, 3	1, 1+, 2	*1, 1+, 2, 2+, 3*
140	3	3	3	3	2, 2+, 3	*3*
160	3	3	3	3	3	*3*

N = 9,944 across 70 languages.

[a] "Typical" refers to a language in which the average proficiency rating falls in the center of the distribution of ratings across all languages. Though language differences exist, these values best represent all of the languages in the dataset. Includes languages not shown.

Analysis of DLPT Version and Maximum Proficiency on Record

To account for potential differences in proficiency ratings due to the rollout of the DLPT 5, we examined the extent to which DLPT test version influenced the maximum DLPT-L and DLPT-R proficiency rating on record for this sample. Specifically, we focused on the maximum proficiency level individuals would be expected to attain if testing on the DLPT 5.

For listening proficiency, the findings showed ratings on the DLPT 5 were lower than those on earlier DLPT versions. Accounting for DLPT version[12] explained an additional 2.7% of the differences in maximum listening proficiency ratings (controlling for DLAB scores). Table C5 (p. 30) shows the maximum listening proficiency ratings SOF operators are expected to attain assuming they are testing on the DLPT 5.

[12] We specifically compared DLPT 5 to prior versions. No comparisons between earlier versions were performed.

There is one notable trend in Table C5 in comparison to the prior analysis, which did not account for DLPT version (Table C3, p. 28):

- **Individuals tended to attain lower levels of listening proficiency on DLPT 5 compared to earlier DLPT versions.** As a result, the maximum proficiency level on record for an individual is expected to be somewhat lower if the rating was produced by a DLPT 5 rather than by an earlier DLPT version for a given language.

Table C5. DLAB Score and Expected Maximum Attained Listening Proficiency (DLPT 5 Only)

DLAB Score	Expected Maximum Listening Proficiency (ILR)					
	Spanish	French	Russian	Chinese-Mandarin	Arabic	*"Typical" Language* [a]
40	1	0+	0+	0, 0+	0	*0+*
60	1	0+, 1	0+, 1	0+	0	*0+, 1*
80	1	0+, 1	0+, 1	0+, 1	0, 0+	*0+, 1*
100	1, 1+, 2	1	1	1	0+, 1	*1*
120	1, 1+, 2, 2+, 3	1	1	1	1	*1*
140	3	1, 1+, 2, 2+, 3	1, 1+, 2, 2+, 3	1, 1+, 2	1	*1, 1+, 2, 2+, 3*
160	3	3	3	2, 2+, 3	1, 1+, 2	*2, 2+, 3*

N = 8,828 across 47 languages.

[a] "Typical" refers to a language in which the average proficiency rating falls in the center of the distribution of ratings across all languages. Though language differences exist, these values best represent all of the languages in the dataset. Includes languages not shown.

The findings for reading proficiency were similar to those for listening proficiency. Reading proficiency ratings on the DLPT 5 were lower than those on earlier DLPT versions. Controlling for DLAB scores, DLPT version explained an additional 3.1% of the differences in maximum reading proficiency ratings. Table C6 (p. 31) shows the maximum reading proficiency ratings individuals are expected to attain assuming they are testing on the DLPT 5. As with listening proficiency, the maximum reading proficiency level on record is expected to be somewhat lower if the rating was produced by a DLPT 5 rather than by an earlier DLPT version.

Table C6. DLAB Score and Expected Maximum Attained Reading Proficiency (DLPT 5 Only)

DLAB Score	Expected Maximum Listening Proficiency (ILR)					
	Spanish	French	Russian	Chinese-Mandarin	Arabic	*"Typical" Language* [a]
40	1, 1+	0	0	0	0	*0*
60	1, 1+, 2	0, 0+, 1	0, 0+, 1	0, 0+, 1	0	*0, 0+, 1*
80	1, 1+, 2, 2+, 3	1	1	1	0	*1*
100	2, 2+, 3	1, 1+	1	1	0, 0+, 1	*1*
120	3	1, 1+, 2, 2+	1, 1+, 2	1, 1+, 2	1	*1, 1+, 2*
140	3	2+, 3	2, 2+, 3	2, 2+, 3	1, 1+, 2	*1+, 2, 2+, 3*
160	3	3	3	3	2, 2+, 3	*3*

N = 8,755 across 43 languages.

[a] "Typical" refers to a language in which the average proficiency rating falls in the center of the distribution of ratings across all languages. Though language differences exist, these values best represent all of the languages in the dataset. Includes languages not shown.

Caveats

The expected maximum proficiency levels presented in Appendix C should be interpreted as the most likely proficiency level individuals are expected to attain based solely on DLAB score using historical norms within the sample population. Individuals with the same DLAB scores testing in the same language will certainly differ in the maximum proficiency levels they attain during their career. The expected maximum proficiency levels presented here should not be interpreted as the maximum *possible* proficiency individuals could attain given unlimited time, resources, and immersion opportunities. The nature of this study (e.g., archival, correlational) does not support that specific interpretation.

While the specific proficiency estimates presented in this study are helpful for visualizing the importance of language aptitude to maximum proficiency, applying them to current or future military personnel assumes no major changes in language training or testing from the historical sample. Two notable and recent changes in military language testing are the introduction of the DLPT Version 5 (i.e., DLPT 5) for certain languages and the move from the DLPT Listening and Reading to the two-skill OPI (speaking and listening) as the ARSOF test of record in 2008. We evaluated the influence of DLPT test version on the proficiency estimates, and present these findings in the Appendix B. The change in proficiency standard from the DLPT to the two-skill OPI in the ARSOF community resulted in a shift in training emphasis towards speaking skills and away from reading skills. This shift in training emphasis may change the maximum proficiencies attained by operators in all language skill modalities. However, it is premature to evaluate the impact the change in testing standard will have on maximum career proficiency for any of the skill modalities examined in this report.

As noted in *Study 1*, IAT training duration for Cat I/II languages was recently increased at USAJFKSWCS from 18 weeks to 24 weeks. This increase in training time for Cat I/II languages is likely to influence the maximum proficiency those training in these languages attain during ARSOF IAT, and

ultimately, their careers in SOF. Additionally, there are other situational factors (e.g., organizational context, testing context) that may influence the maximum proficiency an individual will attain that are not explicitly examined in this study. With the inclusion of both SOF and non-SOF personnel in this analysis, the contexts within which individuals train and test will differ to an even greater extent than would be the case in a SOF-only sample. Therefore, we recommend caution when generalizing the specific proficiency estimates provided in Appendix C to any specific population. However, these variations in the organizational context, testing version, proficiency standard and training duration are not expected to dramatically affect the underlying relationship between language learning aptitude and the acquisition of proficiency, which is the focus of this report. While testing and training practices may alter the levels of proficiency operators attain across skill modalities, those higher in language learning aptitude are still expected to attain greater proficiency than lower aptitude individuals in a comparable testing and training environment.

Conclusions

The objective of this supplemental analysis was to establish the relationship between DLAB scores and the maximum proficiency level military personnel in general attain over the duration of their career. The sample used for this analysis was composed entirely of U.S. Army (SOF and non-SOF) and U.S. Navy personnel. The findings from this analysis show that DLAB scores were predictive of the maximum proficiency individuals in this diverse sample ultimately attained.

As was the case in *Study 2*, it is important to note that the relationships between DLAB scores and proficiency ratings observed in this analysis were small to moderate in magnitude. For a given target language, DLAB scores only predicted between 5% (for speaking) and 11% (for listening) of the differences in operators' maximum proficiency on record. This means DLAB scores failed to predict as much as 89-95% of the differences in maximum proficiency. These differences among SOF operators may be due to other systemic factors (e.g., access and usage of learning resources, opportunity to practice, other trainee characteristics, etc.) that influence proficiency.